

Internal Validity

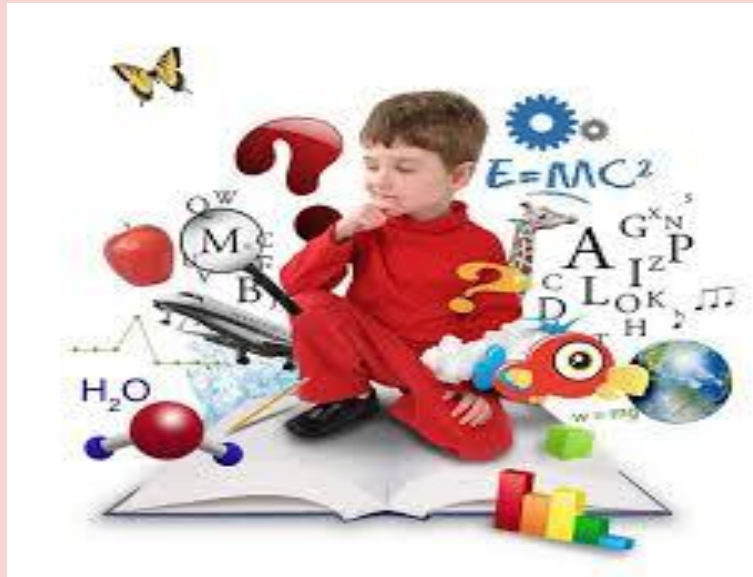
- Has to do with DESIGN and how the constructs are related
- Correlation does not necessarily equal Causation

Ambiguous Temporal Precedence

Confusing the independent and dependent variables

* Are high test scores the result of a good school? Or is a good school the result of good test scores?

Selection



Selection is just what it seems.
You don't select at random, you pick
your own team.
For example, think of the PSSA's...
**If you only measure results of AP
students, is that a reliable way?**

History

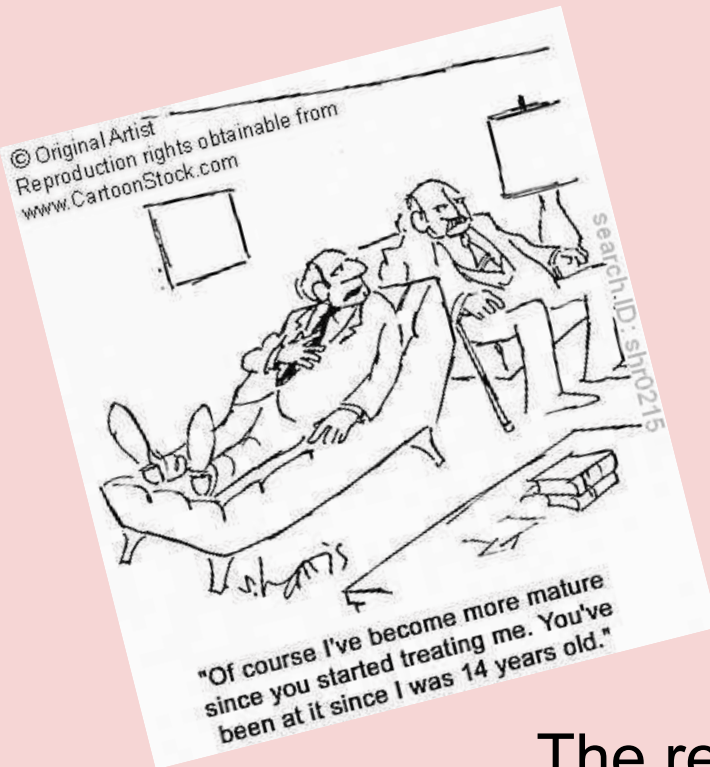
History refers to any external or historical event that occurred during the course of the study that may be responsible for the effects instead of the program itself.

It's no mystery, it is the history
that another event,
could be more prevalent
and destroy our victory!



Sesame and Headstart came out at the same time and researchers are unclear on which one had more of an effect on kids.

Maturation



The research seemed sublime
Until I forgot to calculate time
I thought the results were success
But this aging is a mess
If I'd kept a group in a box
That could have made up for the clocks

Regression

Choosing subjects to show the extreme
 They will likely regress towards the mean
 You should have retested them to show true place
 To reflect a more perfect human race

Cleveland vs. Worcester of A.C. Worcester June 12 1870

FIELDING REC'D.				UMPIRE, <i>Tranley</i>										BATTING RECORD.			
Put Out.	Ass't.	Err's	NAMES, POSITION AND No.	1	2	3	4	5	6	7	8	9	10	At Bat.	Runs	1st B.	Tot B.
4	2	2	1 <i>Dunlap</i> B <i>B-2</i>				<i>H+K</i>			<i>H K</i>							
			2 <i>Hankinson</i> C <i>B-2</i>				<i>A K</i>			<i>H K</i>							
9	1		3 <i>Kennedy</i> 1st <i>S+H</i>				<i>H+K</i>			<i>C+H</i>							
7			4 <i>Phillips</i> A <i>S+H</i>					<i>R-A</i>		<i>H K</i>							
2			5 <i>Skinner</i> R <i>H+K</i>					<i>H K</i>		<i>C-A</i>							
	9		6 <i>McConnic</i> P <i>B-2</i>					<i>H</i>		<i>S+H</i>							
1			7 <i>Geligan</i> M <i>M+H</i>						<i>P-A</i>		<i>R+H</i>						
	2		8 <i>Glasscock</i> S <i>S+H</i>						<i>B-A</i>		<i>H K</i>						
1			9 <i>Hansen</i> L <i>C+H</i>						<i>H K</i>		<i>S-A</i>						
24	14	2	Runs,....	0	0	0	0	0	0	0	0	0	0	27	0	0	
Totals,...				0	0	0	0	0	0	0	0	0	0				
TIME OF GAME: Began, Ended, {				Earned Runs, H. M. 1st Base on Err's,										SCORER:			

1st Strike out by Richmond 5 2 errors by Dunlap in 5th

Attrition:



It happens in a study when members drop out,
They might move, quit, or simply pout.

It's not as rare as it may seem,

They're unwilling to participate despite the
researcher's dream!

Testing

- There is no perfect test.
- Improvements do not always stem from learning.
- Sometimes testing improves testing.
- learn to provide the right answers rather than learn.



Instrumentation



Poor selection of testing instrument impacts the results in skewing of the data (computer based test of writing, when participants lack keyboard skills)

Conclusion Validity

- Has to do with ANALYSIS
 $p =$ between -1 and 1
- Is there an actual relationship between the two variables?

Construct Confounding

Confusing the variables: An additional independent variable, besides the targeted item is affecting the results in the dependent variable

Mono-Method & Mono-Operation Bias

Mono-method Bias - Using just one tool to measure outcomes of a study.

The PSSA reading test does not give you a comprehensive understanding of a child's academic abilities. A better measure would be to include PSSA data along with formative assessments and teacher feedback.

Mono-operation Bias is similar, but related to looking too narrowly at the outcomes.



Like Lays potato chips, you can't use just one...

Mono-style tunnel vision is no fun!

Click Me!



Compensation



Mo Money, Mo Problems—research style

When you do a study, there are lots of complications.
One of them is called **Compensation**!

Control Group Compensation—here's what it means
If you're in the control, you'll probably receive
Money, Sympathy, or TLC
This will affect the validity

Next up we have **General Compensation**,
We'll "bribe you" with rewards for participation.
Depending on what you like, you'll get to choose
We'll give you ipods, money, or shoes.

The point of this is, if you like the reward,
Come see what the study has in store!

Attempting to isolate your testing group from a control group, but unintended conversations and sharing takes place between groups, resulting in a “contamination” of the control group



Treatment Diffusion

Reactivity

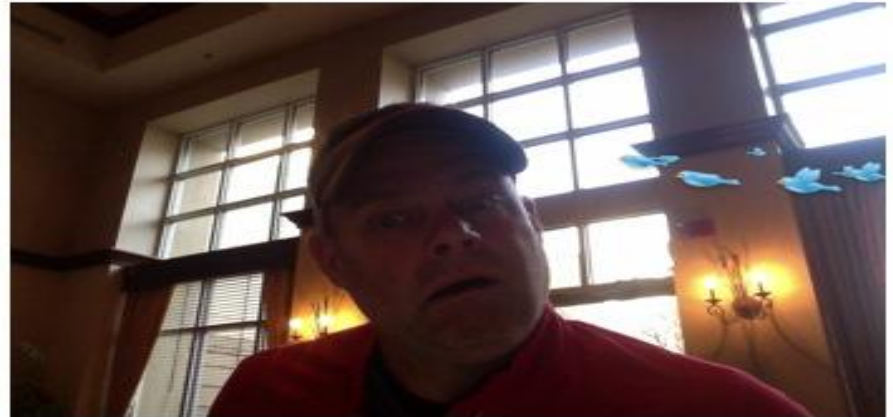
-**Reactivity**: Happens when the subjects responses are affected based on a reaction. The interviewer or an instrument of the study changes or has a significant impact of the metric.

Examples: Pic 1. The interviewer reacts to the subjects responses with a dumbfounded look.

Pic 2: The interviewer responds with laughter

Pic 3: Smiling and appears to be agreeing with the subject

Pic 4: It appears the interviewer is not satisfied with the response.



Construct Validity

- Has to do with having a common agreement upon the **DEFINITION** of the construct
- Did you measure the outcome you wanted to measure?

Low Statistical Power

Do you have enough people in your study?

- * If there are too many, everything is meaningful.
If there are too little, nothing is meaningful.
How many people do you need to count a
“subgroup”?

Fishing and Error Rate

Missing the “needle in the haystack”

- * International test scores... Why did we “lose” to China and Finland? Could it be their school? Their diet? The weather?

Unreliability of Measures

How reliable are your measures over time?

* What happens when you survey someone's recollection about a past event? How reliable are people's opinions over time?

Restriction of Range

There is a ceiling effect which limits results/
outcomes

* High stakes tests are not that difficult for high achieving students - importance of PVAAS over PSSA or Keystone.

Free and Reduced Lunch -
single mother of one child vs.
a family of five with no income.

Heterogeneity of Units

Different result with different groups, so effects
“wash out”

- * different types of strategies work better with
different types of learners

Extraneous Variance in Setting

When you are set up for one expectation - there is something else accounting for your data connection

* Taking tests in rooms you have not learned in, very warm room or a very cold room, losing days to snow

External Validity

- Has to do with SAMPLING and transfer of results
- Are you able to generalize the outcome to the larger population?

lotCR with Units

lotCR with Treatment Variations

lotCR with Outcomes

lotCR with Setting

lotCR with Context-Dependent Mediation